PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 7 : C12Q 1/68 | A2

(11) International Publication Number: **WO 00/08208**

(43) International Publication Date: 17 February 2000 (17.02.00)

(21) International Application Number: PCT/GB99/02579

(22) International Filing Date: 5 August 1999 (05.08.99)

(30) Priority Data:
9817055.8 5 August 1998 (05.08.98) GB

(71) Applicant *(for all designated States except US)*: MEDICAL RESEARCH COUNCIL [GB/GB]; 20 Park Crescent, London W1 4AL (GB).

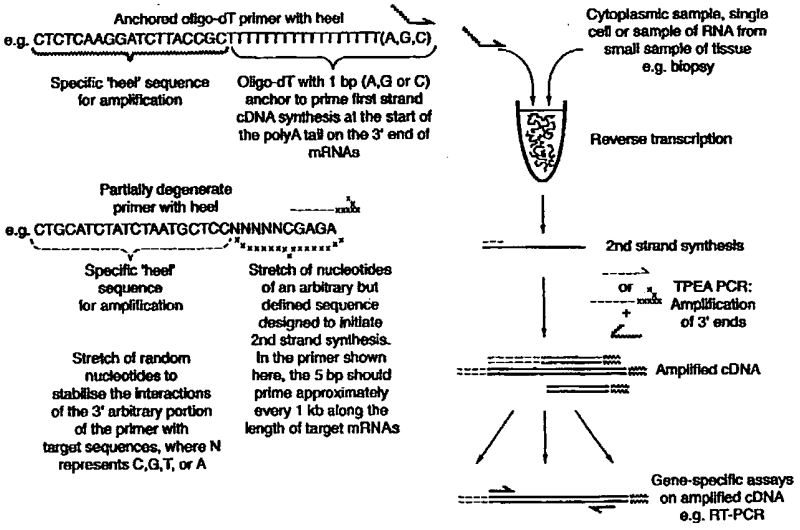
(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: FREEMAN, Thomas, Charles [GB/GB]; 80 Whitehill Road, Cambridge CB5 8LX (GB). RICHARDSON, Peter, John [GB/GB]; Hillcrest Cottage, Bartlow Road, Hadstock CB1 6PF (GB). DIXON, Alistair, K. [GB/GB]; 108 Gwydir Street, Cambridge CB1 2LL (GB).

(74) Agent: HEPWORTH LAWRENCE BRYER & BIZLEY; Merlin House, Falconry Court, Bakers Lane, Epping, Essex CM16 5DQ (GB).

(81) Designated States: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published
Without international search report and to be republished upon receipt of that report.

(54) Title: REVERSE TRANSCRIPTION AND AMPLIFICATION PROCESSES AND PRIMERS THEREFOR



(57) Abstract

A method is provided for the expression profiling of single cells. The method employs a first heeled primer for reverse transcription of mRNA in a sample to provide first strand cDNA species, and then a second heeled primer population to generate second strand cDNAs. The non–heeled portion of the second heeled primers are capable of hybridizing to the reverse transcribed first strands of cDNA species, at least one along the lengths thereof. Due to the presence of random and preselected sequences in the second primers a qualitatively more uniform and therefore representative cDNA profile is produced from cellular mRNAs.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	**ES**	Spain	**LS**	Lesotho	**SI**	Slovenia
AM	Armenia	**FI**	Finland	**LT**	Lithuania	**SK**	Slovakia
AT	Austria	**FR**	France	**LU**	Luxembourg	**SN**	Senegal
AU	Australia	**GA**	Gabon	**LV**	Latvia	**SZ**	Swaziland
AZ	Azerbaijan	**GB**	United Kingdom	**MC**	Monaco	**TD**	Chad
BA	Bosnia and Herzegovina	**GE**	Georgia	**MD**	Republic of Moldova	**TG**	Togo
BB	Barbados	**GH**	Ghana	**MG**	Madagascar	**TJ**	Tajikistan
BE	Belgium	**GN**	Guinea	**MK**	The former Yugoslav Republic of Macedonia	**TM**	Turkmenistan
BF	Burkina Paso	**GR**	Greece			**TR**	Turkey
BG	Bulgaria	**HU**	Hungary	**ML**	Mali	**TT**	Trinidad and Tobago
BJ	Benin	**IE**	Ireland	**MN**	Mongolia	**UA**	Ukraine
BR	Brazil	**IL**	Israel	**MR**	Mauritania	**UG**	Uganda
BY	Belarus	**IS**	Iceland	**MW**	Malawi	**US**	United States of America
CA	Canada	**IT**	Italy	**MX**	Mexico	**UZ**	Uzbekistan
CF	Central African Republic	**JP**	Japan	**NE**	Niger	**VN**	Viet Nam
CG	Congo	**KE**	Kenya	**NL**	Netherlands	**YU**	Yugoslavia
CH	Switzerland	**KG**	Kyrgyzstan	**NO**	Norway	**ZW**	Zimbabwe
CI	Côte d'Ivoire	**KP**	Democratic People's Republic of Korea	**NZ**	New Zealand		
CM	Cameroon			**PL**	Poland		
CN	China	**KR**	Republic of Korea	**PT**	Portugal		
CU	Cuba	**KZ**	Kazakstan	**RO**	Romania		
CZ	Czech Republic	**LC**	Saint Lucia	**RU**	Russian Federation		
DE	Germany	**LI**	Liechtenstein	**SD**	Sudan		
DK	Denmark	**LK**	Sri Lanka	**SE**	Sweden		
EE	Estonia	**LR**	Liberia	**SG**	Singapore		

REVERSE TRANSCRIPTION AND AMPLIFICATION PROCESSES AND PRIMERS THEREFOR

The present invention relates to processes for the reverse transcription of mRNA to provide cDNA. Additionally, the invention relates to processes involving reverse transcription of mRNA and concomitant or subsequent amplification of cDNA. The mRNA samples are typically those obtainable from cell or tissue samples of organisms, such as biopsy or blood samples, for example. In particular, the invention relates to processes for reverse transcription or reverse transcription and amplification of a population of mRNA obtainable from a single cell. Thus, the invention concerns the field of analysis of gene expression (i.e. "expression profiling") of cells and tissues, even down to level of a single cell. The invention also relates to polynucleotide primers adapted for the performance of reverse transcription on population of mRNA species with concomitant or subsequent amplification procedures.

DNA sequence information resulting from genome and expressed sequence tag (EST) sequencing projects is expected to provide the basis for furthering understanding of the control and mode of action of individual gene products. In this respect, expression profiling is regarded as playing a role in the functional characterisation of newly identified genes (Lander, E S (1996), Science, 274:536-539 and Strachan, T (1997), Nature Genetics, 16:126-132).

Many tissues, such as the immune and nervous systems, are composed of highly heterogeneous cell populations. A key factor in understanding their physiology, and the role of specific gene products expressed within them, is to examine gene usage in the context of this cellular diversity. In the past, methods such as Northern blotting and nuclease protection assays were employed to study gene expression. More recently, techniques have been developed for assessing simultaneously the expression of large numbers of genes, e.g. (deRisi, J *et al* (1996), Nature Genetics, 14:457-460; Chee, M *et al* (1996), Science, 274:610-614; Lockhart, D J *et al* (1996), Nature Biotech, 14:1675-1680; Madden S L *et al* (1997), Oncogene, 15:1079-1085; and Marshall A & Hodgson J (1998), Nature Biotechnology, 16:27-31). All these techniques, however, require relatively large amounts of RNA and currently lack the sensitivity to analyse specimens derived from small populations of cells or indeed from an individual cell.

At present, methods for the analysis of gene expression within single cells or small tissue samples are limiting. Whilst *in situ* hybridization techniques provide detailed information about the cellular expression pattern of a gene in intact tissue, be it whole-mounts or tissue sections, the technique is relatively laborious and unable to analyse multiple transcripts in a single preparation.

In recent years, the polymerase chain reaction (PCR) has been used successfully to investigate gene

expression in cytoplasmic samples derived from single cells. The nested-primer approach has a good level of sensitivity, but the analysis is restricted to just a small number of closely related genes from specific gene families (Lambolez, B *et al* (1992) Neuron, 9:247-258; Sucher, N J *et al* (1993), J Biol Chem, 268:22299-22304; and Yan, Z & Surmeier, D.J (1997), Neuron, 19:1115-1126).

Other techniques known to be capable of detecting the expression of unrelated genes in a single cell include T7 RNA polymerase amplification of mRNA (Eberwine, J *et al* (1992) Proc Natl Acad Sci USA, 89:3010-3014; Van Gelder *et al* (1990), Proc Natl Acad Sci USA, 87:1663-1667). Also known are methods employing PCR after prior homopolymeric tailing of the first strand cDNA (Brady G & Iscove, N.N. (1993), Methods in Enzymol, 225:611-623; Frohman *et al* (1988), Proc Natl Acad Sci USA, 85:8998-9002; and Jena P K *et al*, (1996), J Immunol Methods, 190:199-213). However, neither of the aforementioned approaches have been demonstrated to be capable of analysing more than a small number of genes of the total number of genes expressed in any given cell. Moreover, the techniques are not widely used. The former procedures suffer from technical difficulties, whilst the latter procedures are biased against long transcripts and often requires subsequent cloning of the amplified products. In view of the known technologies, there is still a great need for improved methods which can quickly, easily, sensitively and simultaneously reveal for analysis the spectrum of gene expression at the cellular level, even right down to the level of a single cell.

Significantly, there is also a need for understanding gene expression at the level of single cells and small biopsy samples. A method capable of assessing gene expression in samples as small as a single cell would be of considerable benefit to the understanding of gene expression and so the molecular basis of cell function, the identification of specific types of diseased (e.g. cancerous) cells and in the future routine analysis of gene expression in the scientific and wider communities. (With the increasing influence of genomics on Biomedical Science, the ability of scientists and medical doctors to assess gene expression in small tissue samples will be of increasing utility.)

Currently, there is also considerable interest in the use of multiple gene arrays to assess the expression in tissue samples of many hundreds or thousands of genes. However, although this approach allows many genes to be examined simultaneously in a single sample, relatively large amounts of starting material are required. Thus, any data obtained cannot be unequivocally assigned to individual cell types in an original sample and because of this the samples have to be large (requiring more than 10^7 cells). This causes considerable problems in the collection of samples from humans and in the interpretation of the data. This problem is particularly acute in the CNS and immune systems where the heterogeneity of the cells means that such data is extremely difficult to interpret.

Previous attempts at assessing gene expression at the level of single cells has been of only limited success. The following are presented as examples of known techniques.

Nested PCR

Nested PCR has been used successfully in many scientific laboratories and it relies on two sequential amplification steps, both targeted to the genes of interest. This means, therefore that in any one cell sample the expression of only a few genes (of up to 3 gene families) can be examined. The technique does not ensure that all members of cDNA populations in a complex mixture are amplified, nor that all the amplicons are of similar sizes. The technique suffers from the drawback that only a few gene families can be examined at a time. This is wholly unsatisfactory from the point of view of expression profiling.

RNA Amplification

This is a technically complex procedure and its manifest difficulties are seen in the few laboratories in the world that are able to use this technique regularly. The technique is just not suited to a simultaneous analysis of a large number of cell samples.

In Situ Hybridization

This technique is suited to the analysis of the expression of no more than about 3 genes in any cell.

The technique is useful for simultaneous expression screening of a large number of dead cells in fixed tissue slices *In situ* hybridization usually involves the use of radiolabels which are inconvenient and the technique as a whole is quite unsuitable for expression profiling of living cells.

cDNA Tailing

This technique used by a number of research groups employs terminal deoxynucleotide transferase to attach a unique priming site on to the 3' end of a first strand cDNA following reverse transcription. Following PCR of the cDNA with the unique priming site, the expression profile of a number genes from a single cell can be analysed. The technique has a number of drawbacks. First, there is the need to use homopolymeric PCR primers capable of annealing to sites in a DNA sequence. Also, there is an unequal amplification of cDNA because unequal lengths of the cDNA transcripts are amplified. The amplification efficiency is low. The initial PCR reactions for the different transcripts operate at vastly different efficiencies, and so bias the procedure in favour of shorter gene transcripts.

Ligation

In this technique, a primer sequence is ligated to the 3' end of cDNA to provide a second amplification primer site. However, this technique suffers from the same problems as the cDNA tailing technique referred to above. PCR of the different cDNA species in the reverse transcribed sample takes place at

different levels of efficiency, depending on the length of the cDNA molecule being amplified. Additionally, the ligation reactions can be difficult to control with multiple priming sites being ligated.

Each of the aforementioned techniques suffers from a variety of limitations. Indeed, an analysis of the scientific literature suggests that none of these techniques (other than *in situ* hybridization) are widely used in practice. Existing methods for analysis of gene expression in small samples of RNA, particularly from single cells, are severely limited in terms of the number and diversity of genes that it is possible to analyse, and the difficulty of experimental procedures involved.

The present inventors have sought to develop a more straightforward, reproducible and reliable cDNA amplification procedure for small mRNA samples wherein expression profiling can be conducted, thereby avoiding the various problems outlined above.

Accordingly, in first aspect the present invention provides a process of reverse transcribing mRNA species present in a sample from an organism comprising:

- reverse transcribing the mRNA species using a first heeled primer, thereby to provide first strand cDNA species;

- synthesising second cDNA strands using a second heeled primer population, the non-heel portion of the second primers being capable of hybridizing to

the reverse transcribed first strand cDNA species at least once along the lengths thereof.

In second aspect the invention provides a process of reverse transcribing and amplifying mRNA species present in a sample from an organism comprising:

- reverse transcribing the mRNA species using a first heeled primer, thereby to provide first strand cDNA species;

- synthesising second cDNA strands using a second heeled primer population, the non-heel portion of the second heeled primers being capable of hybridizing to the reverse transcribed first strand cDNA species at least once along the lengths thereof;

- amplifying the resulting cDNA species.

In other words, in both of the first and second aspects of the invention described above, the particular sequence of the non-heel portion of each second primer in the second heeled primer population is such that at least one second heeled primer from the population is capable of hybridizing to each reverse transcribed first strand cDNA species. For each first strand cDNA species, the number and identity of individual second heeled primers which hybridize thereto is expected to be different but this is not an expectation which excludes the possibility of similarities in hybridizations with cDNAs arising. Furthermore, due to the sequence characters of the non-heel sequences of the second

primers, their hybridization with the first cDNA strands is possible at a multiplicity of sufficiently complementary sites along the lengths of the cDNA strands.

The particular temperatures, enzymes and reagents (other than the first primer) used in the process of reverse transcription may be those already known in the art.

A "heeled" primer will be readily understood in the art to be a primer comprising a hybridizing portion and a non-hybridizing portion, wherein the non-hybridizing portion represents the "heel" of the primer.

The second primer is actually a population of individual primer species. When the first strand cDNA population is contacted with the second primer population under appropriate hybridizing conditions then because of the selection of nucleotide sequences amongst the second primers, each cDNA species will hybridize with at least one second primer, second cDNA strand synthesis then proceeds in a 5' to 3' direction from the hybridized second primer. Although the inventors do not wish to be bound by any particular theory, what appears to happen is that when more than one second primer species hybridizes to any given cDNA species then second strand cDNA synthesis proceeds in its 5' to 3' direction from the second primer whose 3' end is not obstructed along the first strand cDNA template by any other second primer hybridized thereto. As a result, and particularly where a multiplicity of second primers

hybridize to the first strand cDNA template, there is more of a tendency to generate second strand cDNA molecules starting at points further upstream on the first strand cDNA template, i.e. further downstream on the original mRNA molecules. Coincidentally, the 3' end portions of mRNA molecules are generally more diverse in their sequences.

A net result of the reverse transcription process of the invention is that there appears to be a bias towards a more uniform length of cDNA molecule. This in turn impacts on any subsequent amplification procedure. When the amplification is PCR for example then more uniform length cDNA molecules lend themselves more to amplification than a population of cDNA molecules less uniform in length.

Advantageously, the processes of the invention generate cDNA molecules highly representative of the spectrum of mRNA molecules in a sample. The element of bias towards more uniform length cDNA molecules ensures that even relatively low abundance mRNA species are transcribed, and optionally amplified, to the same level of efficiency as more abundant mRNA species. Thus, a much better qualitative profile of expressed gene sequences in samples can be achieved than was hitherto possible.

The degree of sensitivity of the processes of this invention can of course be varied by modifying the numbers and sequences of the second primer population species thereby modifying the frequency with which the second primers hybridize to the first strand cDNA templates.

Where further amplification of the cDNA products is required, the amplification of the cDNA species resulting from the reverse transcription preferably employs a third primer comprising at least a part of the heel portion of the first heeled primer and a fourth primer comprising at least part of the heel portion of the second heeled primer. Further amplification may be advantageous where subsequent analysis of cDNA species involves less sensitive detection means or where a larger sample is required for analysis by methods which require larger quantities of cDNA material.

The second heeled primer population may comprise primers differing by up to five nucleotide bases the population preferably comprising a number of primers in the range 1000 to 100,000 primers, more preferably in the range 1024 to 65536 primers. In order to achieve this, the primers of the second heeled primer population preferably each comprise a random sequence of nucleotides in the range of 5 to 8 nucleotides 3' to the heel and a further sequence of at least 5 nucleotides contiguous 3' therewith. As will be appreciated, where there are 5 random nucleotides (which is preferred) there will be 4^5 (i.e. 1024) possible pentamer sequences.

The further sequence of nucleotides may be selected by sequence analysis of known sequences so as to promote the ability of the second heeled primer as a whole to hybridize to the transcribed cDNA species. The sequence analysis can be carried out through databases of DNA or RNA sequences. In particular,

the known sequences of the organism of interest are preferably consulted. The further sequence of nucleotides preferably comprises a number of nucleotides in the range 2 to 10 nucleotides. In a particularly preferred embodiment the further sequence of nucleotides may comprise a number of nucleotides equivalent to the number of nucleotides in the random sequence of nucleotides.

The further nucleotide sequence of the second heeled primers is preferably constant throughout the population of these primers and it is selected so as to stabilise the primers and to ensure optimal efficiency of hybridization to target first strand cDNA species.

In preferred embodiments, the second heeled primer from the population of second primers preferably hybridises on average once in every 1kb portion of first strand cDNA species. This has been found to provide a relatively efficient and uniform reverse transcription and optionally amplification of mRNAs in a sample.

A particularly preferred further sequence of nucleotides in the second primers is:

CGAGA

and a particularly preferred second heeled primer population is represented by:

CTGCATCTATCTAATGCTCCNNNNNCGAGA

wherein N is independently selected from C G T or A.

The heel portion of the first and second heeled primers are preferably selected so that they lack the ability to hybridise to mRNA or first strand cDNA respectively. The heel portions, like the further sequence portions of the second primers, are selected by an analysis of known nucleotide sequence information. In particularly preferred embodiments, the heel portions preferably comprise sequences absent from the mRNA species in the sample, although the heel portions may simply comprise sequences absent from the genome of the organism from which the sample is taken. The heel portions preferably comprise a number of nucleotides in the range 15 to 50, more preferably 18 to 22 nucleotides although somewhat fewer or somewhat more nucleotides may be acceptable.

In preferred embodiments, the first heeled primer is preferably an anchored primer comprising an oligo (dT) sequence. The nature of anchored primers is already well known in this art. For example, provision of a few non-T bases at the 5' end of the primer ensures that hybridization of the primer occurs at the 5' end of the mRNA poly A tail.

The fourth primer is preferably the heel of the second heeled primer, or at least a portion thereof.

The third primer is preferably the heel of the first heeled primer, or at least a portion thereof.

The third primer may be the same as the first heeled primer and this can be advantageous in reducing the numbers of reagents needed to perform the processes of the invention.

The frequency with which individual second primer population species hybridize along a given length of nucleic acid may be adjusted by employing suitable hybridizing conditions. Preferably, the hybridization conditions are of limited stringency so that the random sequences of oligonucleotides in the second primers have a significant effect on whether hybridization occurs or not. The degree of stringency of hybridizing conditions and the number of contiguous random bases in the second primers may be varied according to routine trial and error in order to achieve a desired frequency of hybridization of second primer species along a given length of nucleic acid material.

The amplification of the resulting cDNA species preferably comprises more than one round of amplification cycles. Preferably, each further round of amplification comprises addition of further second and third primers and amplification reagents, optionally a fourth primer as well.

Each round of amplification may comprise 5 to 45 cycles, more preferably 10 to 40 cycles. In particular, the first round of amplification may comprise a lesser number of cycles than any further rounds of amplification.

The amplification process is preferably PCR although modified PCR procedures or other compatible amplification procedures may be used.

Preferred PCR cycles comprise X_1°C for Y_1 min; then X_2°C for Y_2 min; then X_3°C extension for Y_3 minute; then Y_4 min extension, wherein X_1, $>X_3$ $>X_2$ and Y_4 $>Y_3$ $>Y_2$ $>Y_1$. X_1 is in the range 90 to 94°C, X_2 is in the range 45 to 70°C, X_3 is in the range 65 to 75°C and Y_1, Y_2, Y_3 are in the range 15 seconds to 4 minutes and Y_4 is in the range 2.5 to 20 minutes. Particularly preferred values are X_1 = 92°C, X_2 = 60°C, X_3 = 72°C, Y_1 = 2.5 minutes, Y_2 = 1.5 minutes, Y_3 = 1 minute and Y_4 = 10 minutes.

The sample from an organism will preferably include or be derived from tissue or cells. The sample may be comprised of whole cells, possibly comprising a single cell type, even comprising just a single cell. In particularly preferred embodiments the samples are composed of the cytoplasm of cells (same cell type or mixture) or more preferably the cytoplasm of a single cell. Biopsy samples may provide useful sources of sample material ranging from a few grams to a few micrograms of tissue/cell material.

The cytoplasm may be obtained by lysis or aspiration of a cell or cells and such cell may be obtained by fluorescence activated cell sorting (FACS). The processes of the invention are sufficiently reliable, sensitive and efficient that substantially all mRNA species in a sample are reverse transcribed and optionally amplified to approximately the same degree. A more accurate and reliable picture of gene

expression can be obtained for a cell sample. Advantageously, the processes of the invention permit single cell gene profiling.

In third aspect the invention provides a method of reverse transcribing expressed gene sequences in a sample from an organism comprising reverse transcribing the mRNA in the sample using a first primer to produce first strand cDNA species, synthesising second cDNA strands using a population of second primers, wherein at least one second primer in the population hybridises to a given first strand cDNA species. This method may further comprise the amplification of the resulting double stranded cDNA. Preferred or alternative versions of this method may comprise one or more of the further features of the other processes of the invention as hereinbefore described.

In fourth aspect the invention provides a polynucleotide primer for reverse transcription of mRNA species comprising an oligo (dT) sequence and 5' thereto a polynucleotide heel sequence, wherein the heel sequence is substantially incapable of hybridisation to the mRNA species. The primer is preferably an anchored primer and may comprise the further features as hereinbefore described.

In fifth aspect the invention provides a polynucleotide primer population for synthesis of second strand cDNA species from first strand cDNA species, wherein at least one primer in the population is capable of hybridising to a given first strand of cDNA. At least one primer in this primer

population is preferably capable of hybridising approximately at least once in any given 1kb of first strand cDNA. The primers may further comprise one or more additional features of such primers as hereinbefore described.

In sixth aspect the invention provides polynucleotide primers for amplification of cDNA comprising a reverse transcription primer (ie the first primer) as hereinbefore described and a primer comprising at least a portion of the heel portion of the second primer population as hereinbefore described.

In seventh aspect, the invention provides the use of a polynucleotide comprising an oligo (dT) sequence and a heel sequence 5' thereto for the reverse transcription of mRNA species in a sample. Such a polynucleotide may further comprise one or more of the characteristics of the first primer as hereinbefore described.

In eighth aspect the invention provides the use of a polynucleotide primer population as hereinbefore defined for the synthesis of second strand cDNA from a population of first strand cDNA species.

In ninth aspect the invention provides a cDNA library preparation obtainable by a process or method as hereinbefore defined, said library comprising substantially all cDNA species corresponding to genes expressed by a single cell, cell type or tissue.

In tenth aspect the invention provides a kit for the production of cDNA from mRNA in a sample from an

organism comprising a primer of the fourth aspect of the invention and a primer population of the fifth aspect of the invention. The kit may further comprise at least one further primer for achieving amplification of the cDNA. Particularly preferred kits comprise a primer pair for amplification of the cDNA.

The invention thus provides a rapid, robust and reproducible procedure, called Three Prime End Amplification (TPEA), optionally with PCR (TPEA-PCR), capable of amplifying 3' fragments of cDNA prior to analysis by other techniques. An important advantage of TPEA-PCR, is the relative ease of performing the method. Other known procedures are generally time consuming and complex, involving DNA purification and precipitation from one step to another. The present cDNA amplification technique however, can be carried out in a single tube with a need for only limited manual intervention. This therefore makes it possible to amplify large numbers of samples relatively easily. The ability to then analyse the expression of many genes of unrelated sequence, both at high and low abundance, in samples taken from as little as a single cell, will potentially allow it to be used in high throughput screening systems.

The invention can be used to analyse gene expression in samples as small as just a single cell (Figure 4), or in much larger samples such as 100 cells (Figure 2). Amplification from a single cell currently provides enough material for approximately 40 gene specific PCR reactions. Whilst this is already an improvement over existing protocols, it should

theoretically be possible to improve the efficiency of the TPEA reaction to provide far higher yields of 3' cDNA product. This would then not only allow the number of gene-specific PCR reactions performed on each sample to be increased, but more importantly allow the procedure to be linked to other analysis procedures. With improvements to the amplification regime and addition of fluorescent or radioactive nucleotide label to the reaction, it should be possible in the future to analyse the product using array based hybridization technologies (deRisi J *et al* (1996), Nature Genetics,14:457-460, Chee M *et al* (1996), Science, 274:610-614, Lockhart D J *et al* (1996), Nature Biotech, 14:1675-1680) which currently require relatively large amounts of RNA for a single assay. Such developments would potentially allow the expression profiling of hundreds or thousands of genes in samples derived from biopsies or single cells.

The inventors provide a rapid, robust and reproducible procedure, called Three Prime End Amplification optionally with PCR (TPEA-PCR) (TPEA), capable of amplifying 3' fragments of cDNA prior to analysis by other techniques. In the method of the invention, the 3' region of mRNA is amplified arbitrarily by PCR using a combination of primers. The amplified cDNA, which represents the most diverse region of gene sequence, can then be analysed by a second round of PCR using gene-specific primers. Using the invention it is possible to analyse the expression of, for example, up to 40 genes (20 in duplicate) in single human lymphoblastoma cells. The method is also suited to the analysis of genes

expressed at low levels in small populations of cells, eg expression of the adenosine A_{2a} receptor in cholinergic neurons of the rat striatum.

Sequence diversity between genes is at its greatest in the 3' untranslated region and this region provides the most unique target for gene-specific assays; this is especially important when wishing to differentiate between closely related members of a gene family. The procedures of the invention preferentially amplifies this portion of the mRNA sequences. cDNA synthesis by reverse transcriptase is initiated by an anchored oligo-dT priming so that the 3' region of all genes is represented in the resulting single-stranded cDNA. A 5'-specific heel may be incorporated into this primer for use in the subsequent amplification procedure. For second-strand synthesis, it is desireable that about 1 kb from the 3' end of each gene is selected and amplified by PCR Assuming a completely random length of nucleotide sequence, it would be expected that a given 5 base sequence would appear every 1024 bases (4^5), even though some nucleotide sequences are more common than others (Lopez-Nieto & Nigam (1996) Nature Biotech 14:857-861). A pentameric sequence is preferentially selected so that the primer initiates second-strand synthesis in an arbitrary manner within 1 kb of the 3' end of the mRNA. A search of 30 gene sequences reveals that at least one copy of this 5 base sequence was present in this region of each gene. 5' to this, 5 bases of random sequence (N_5) are preferably incorporated in order to stabilise the interaction of the arbitrary pentameric sequence, which in turn was flanked by a specific heel

sequence. After a single round of second-strand synthesis, each DNA strand contains a specific priming site 5' and 3' to the region of interest, thus allowing amplification of the intervening sequence. The majority of the mRNA species represented in the first-strand cDNA pool before amplification, as detected by conventional RT-PCR are also detectable after amplification.

The invention should permit the analysis of gene expression in samples obtained from small samples of tissue or single cells. In so doing, it should allow the utilisation of the wealth of new sequence data now available, to further understanding of disease processes and the cellular physiology of complex issues.

A major utility of TPEA-PCR will be in sampling single cells in tissues and in culture conditions to conduct detailed studies of temporal gene expression, changes in gene expression in response to growth conditions or environmental insults, and in identifying hitherto undetected gene activity associated with particular cellular events. Apart from the study of cells to understand their innate biology and responses, new approaches to toxicology profiling are promised as well as means to molecularly classify rare cell types.

The invention has many possible applications including

1. Making a single cell cDNA libraries for subsequent detailed analysis of gene expression, and the discovery of novel genes.
2. Real time profiles of gene expression in selected cells.

3. Preamplification of small (and single cell) samples for subsequent analysis of gene expression using hybridization based assays including those using cDNA and oligonucleotide arrays.

4. Analysis of gene expression in small tissue samples (diagnosis of cancer and other disease states).

5. Analysis of gene expression by PCR.

6. Amplification of full length RNA samples from single cells and small samples, for subsequent library making or expresion in expression systems.

Preferred embodiments of the invention will now be described in more detail by way of specific examples and drawings in which:

Figure 1 shows a schematic illustration of the process of TPEA-PCR. Details of the protocol are given in the Examples, as are the particular sequences of the anchored oligo (dT) primer and the partially degenerate second strand primer.

Figure 2 shows cDNA amplification. cDNA was prepared from varying numbers of sorted lymphoblastoma cells, and analysed by RT-PCR for the expression of 8 "housekeeping" genes, before (left) and after (right) TPEA-PCR. After cDNA amplification, expression of all 8 genes assessed could be detected after carrying out gene specific PCR on only 5% of the amplified cDNA generated from a single cell. Without cDNA amplification however, expression of none of the genes assayed could be detected at this level. Genes assayed: RPL21 (riboprotein L21), RP27a (riboprotein 27a), RPL28 (riboprotein L28), RPS5 (riboprotein S5), HSKPQZ7 (housekeeping protein), ACTB (Cytoplasmic beta-actin), G-3-PDH (Glyceraldehyde-3-phosphate dehydrogenase) and EF1 (Elongation factor 1).

Figure 3 shows multiple gene expression analysis in single lymphoblastoma cells. Four cells (1-4) were lysed, reverse transcribed, the cDNA amplified and gene specific PCR performed on the product. CD2, SI and intron primer pairs serve as negative controls to check for genomic contamination of the samples. Eleven of the other genes are expressed in all four cells in duplicate, while JUND, EF1 CDC25B and CD19 expression was not consistent between cells. Genes assayed: RPL5 (riboprotein L5), RPL21 (riboprotein L21), RP27a (riboprotein 27a), RPL28 (riboprotein L28), RPS5 (riboprotein S5), RPS9 (riboprotein S9), RPSI0 (riboprotein SI0), RPS29 (riboprotein S29), HSKPQZ7 (Housekeeping protein), ACTB (Cytoplasmic beta-Actin), G-3-PDH (Glyceraldehyde-3-phosphate dehydrogenase), EF1 (Elongation factor 1), JUND (JUND), CDC25B (cell cycle factor CDC25b) and the cell surface antigens CD19, CD79a, CD2, IGM

(immunoglobulin IgM), SI (the intestine-specific enzyme, sucrase-isomaltase).

Figure 4 shows adenosine 2a receptor expression in striatal cholinergic interneurons. Panel a shows an infrared video image of a rat striatal cholinergic interneuron during electrophysiological characterisation and panel b, after aspiration of cytoplasm. The expression of four housekeeping genes, the transmitter synthesising enzymes choline acetyltransferase (found only in cholinergic neurons) and glutamic acid decarboxylase (found in GABAergic, medium spiny neurons), three tachykinin (NK) receptors and the adenosine A_{2a} receptor was assessed in twenty six striatal cholinergic neurons. Two representative neurons are shown, neuron 1 expresses the A_{2a} receptor, neuron 2 does not, while the expression of the other genes tested are the same in both.

Example I - TPEA-PCR of Lymphoblastoma Cells

TPEA-PCR assay was performed on lymphoblastoma cells in the G0/G1 phase of the cell cycle. Groups of 100, 10 and single cells were flow sorted into wells containing lysis buffer and the mRNA reverse transcribed. A proportion of the sorted cells then underwent 3' end amplification, as described hereinafter. Figure 1 shows a schematic summary of the TPEA-PCR procedure. Gene-specific PCR assays for 8 'housekeeping' genes were carried out on lymphoblastoma cDNA, before and after cDNA amplification. Following reverse transcription only, the expression of each of the housekeeping genes

could be detected when cDNA generated from between 1 and 100 cells was used in each PCR assay as shown in Figure 2. It was not however possible to detect the expression of any of these genes when cDNA equivalent to less than one cell was assayed. Following TPEA-PCR however, the expression of all eight of the genes was detectable even when amplified cDNA generated from the equivalent of as little as 5% of one cell was present in the PCR assay (Figure 2).

Lymphoblastoma Cell Sorting

An Epstein Barr virus transformed lymphoblastoid cell line (HRC575, ECACC, Porton Down, UK) was maintained in log phase growth in RPMI 1640 medium supplemented with 16% foetal calf serum, 2mM L-glutamine and penicillin-streptomycin (100 U/ml and 100 mg/ml respectively). Cells at approximately 10^8 per ml were stained with the bixbenzimadazole dye Hoechsht 33342 (Sigma, Poole, UK) at 1 µg/ml for 30 minutes at 37°C. Cells were sorted by using the Autoclone attachment of a Coulter Elite ESP flow cytometer, 300 mW of all lines UV from a Coherent 306 laser and by using single drop and complete abort sorting settings. Time of flight, forward and right angle scatter, and Hoechst fluorescence peak and area measurements were used to ensure the sorting of single cells. The accuracy of sorting (both spatial and numerical) was tested by sorting single fluorescent beads (DNA Check, Coulter Corp) into 96 well plates and viewing the plates on a fluorescence microscope.

Reverse Transcription (RT) and cDNA Amplification

Lymphoblastoma cells were FACS sorted into 96 well plates containing 7 µl of freshly prepared lysis buffer (50 mM Tris-HC (pH 8.3), 75 mM KCl, 3mM $MgCl_2$, 5mM NP-40 (Sigma) and 1.5 units of RNase inhibitor (Pharmacia, Milton Keynes, UK). This buffer leaves the nucleus intact (Jena *et al* (1996), J Immunol Methods, 190:199-213). After 5 min on ice, nuclei were removed by centrifugation (8,000g, 5 min at 4°C), and the supernatant aspirated and the RNA reverse transcribed in a reaction volume of 10 µl containing; 1 x first-strand buffer, 200 Units M-MLV reverse transcriptase (Gibco BRL, Paisley, UK), 0.5 ng reverse transcription primer for 60 min at 37°C. The RT primer was composed of an anchored oligo(dT) primer with a specific 5' heel sequence: CTCTCAAGGATCTTACCGCTTTTTTTTTTTTTTT (A,G,C).

Second-strand cDNA synthesis was initiated by incubation of the first-strand cDNA with 1 ng of a primer consisting of (5' to 3'); a 20 base sequence selected due to its absence from the mammalian databases, a stretch of five random nucleotides and a defined pentameric sequence (CTGCATCTATCTAATGCTCCNNNNNCGAGA where N represents C, G, T or A) for 15 mins at 50°C under amplification conditions described below. Although this primer will undoubtedly prime second-strand DNA synthesis at many sites on the first strand cDNA, the subsequent PCR between the heel sequence of the oligo(dT) primer and the arbitrary primer closest to the 5' end, ensures amplification of cDNA sequences complementary to the 3' ends of the polyA tail. After allowing the second-strand primer to anneal, primer extension was

performed at 72°C for 10 min using AmpliTaq DNA polymerase (0.35 units, Applied Biosystems, Warrington, UK) in PCR buffer containing 67 mM Tris HCl (pH 8.3), 4.5 mM $MgCl_2$ and 0.5 mM dNTPs. Subsequently, 0.4 ng of 3' heel primer (CTCTCAAGGATCTTACCGC) was added and the reaction subjected to 10 cycles of 92°C for 2.5 min, 60°C for 1.5 min and 72°C extension of 1 min, followed by a final 10 min extension. A further 125 ng of second-strand primer and 50 ng of 3' heel primer were then added in 10 µl of PCR reaction mix. After 15 cycles (as before), a further 10 µl of PCR reaction mix containing 125 ng second-strand primer and 50 ng of 3' heel primer were added to the reaction and subjected to another 15 rounds of PCR. The final reaction mixture was then diluted to 200 µl with 10 mM Tris/0.1 mM EDTA (pH 8.1). 5 µl samples used for subsequent gene specific PCR assays.

Gene-specific PCR

Samples (5 µl) of amplified cDNA were subjected to hot-lid PCR carried out in 1 x PCR buffer (3.5 mM $MgCl_2$ pH 8.8) containing, 12.5% sucrose, 0.1 mM cresol red, 12 mM β-mercaptoethanol, 0.5 mM dNTPs (Pharmacia), 0.6 U AmpliTaq DNA polymerase (Applied Biosystems), and primers were used at 100 ng/reaction. Amplifications were carried out on PTC-225 thermal cyclers (Tetrad, MJ Research, US). Following an initial 2 min denaturing step (92°C), each PCR cycle consisted of 30 sec denaturing (92°C), 90 sec annealing (55°C), and 60 sec elongation (72°C). After the final cycle the reaction was held

for 10 min at 72°C. The PCR products were then separated on a 2.5% agarose gel, stained with ethidium bromide and photographed. All gene-specific primers are listed in table 1 set out below.

Table 1. Primer Sequences used for gene specific assays

Human Primers

Gene	Symbol	Acc. No.	Forward	Reverse
Riboprotein 1.5	RPL5	U14966	GAACAGCGTAACTCCAGACATG	CCGCTCCTGAGCTCTGAG
Riboprotein L21	RPL21	U14967	GTGCGTATTGAGCACATAAAGC	TTCCAGCAGCTCAGGCTC
Riboprotein L27a	RPL27a	U14968	GACAAATTGTGGACTTTGGTCA	ATGTGCTTCAAGCCACCAG
Riboprotein L28	RPL28	U14969	ATCAACAAGAATGCTCGCG	TGACTTTATTGCTCTGGGGG
Riboprotein S5	RPS5	U14970	GTGAACCAGGCCATCTGG	CCAAAGGGCAGACAGGTTTA
Riboprotein S9	RPS9	U14971	CATTGTCCGCCTGGATTC	TGTTTATTTGGCAGGAAAACG
Riboprotein S10	RPS10	U14972	GGAGATTGTGCCTGCCAC	AAACTGGAATTCGGTTGCTG
Riboprotein S29	RPS29	U14973	ATTCGGCCAGGGTTCTC	TTTTTCATTGAGTAGATGCCCC
Housekeeping protein Q27	HSKPQZ7	M81806	GGAACTTCCTCTGGGAACCT	GGAGGTCAAGTCAAGCTCCA
Glyceraldehyde-3-Phosphate Dehydrogenese	G-3-PDH	M33197	CGACCACTTTGTCAAGCTCA	AGGGGTCTACATGGCAACTG
Cytoplasmic beta-actin	ACTB	AB004047	CGTGGACATCCGTAAAGACC	ACATCTGCTGGAAGGTGGAC
JUND	JUND	X56681	CTCGGGGAACAAACGTTG	ACAAACAGGAATGTGGACTCG
Elongation factor 1	EF1	M82882	CCAAGAAAGCAACTTGAGCC	TGCAGAAGACATTTTTTTCTTG
Cell division cycle 25B	CDC25B	M81934	AGTGCCTTGCATACCCAAAC	CCCCCATCTTGAGATTCAGA
CD19	CD19	X13312	CCAACCTCTGGAGCAATGTT	GGAATACAAAGGGGACTGGA
CD79A	CD79A	L32754	CTTCTGGGGGCTTCCTTAGT	GTTAGGAGGTGGGGCAGTTT3
Immunoglobulin M	IGM	X17115	GTCCCCGGAGAAGTATGTGA	GACTTGTCCACGGTCCTCTC
CD2	CD2	M16445	TCTTCGAACTCAGCCATGTG	GGCTGCTTGTAGTGAGACCC
Sucrase-isomaltase	SI	X63597	CAGGGTTCTCTGTTTTGGGA	GGAAGCGAATTTTTATTTCCG
INTRON			GTGGATATAGTTGAGTCC	CACTGTTGACCACTCCAT

Rat Primers

Gene	Symbol	Acc. No.	Forward	Reverse
Beta-actin	Aclb	V01217	CATCGATGCCCTGAGTCC	ACACCTCAAACCACACTCCCAG
Alpha-tubulin	Tuba	V01226	CAGTGGTACGTGGGTGAGG	TTTGACATGATACAGGGACTGC
Cyclophilin	Cyca	M19533	ACTGCCAAGACTGAGTGGCT	AATGGTTTGATGGGTAAAATGC
Ribosomal protein L18	Rpl18	M20156	AGGTGTACCGACACTTTGGC	TTTTGTAGCCTCGGCTGG
Choline acetyllransferase	Cha1	*	TACTAAGCTCTGTTCCCATCCC	ACCCAGGTTGCTTCCAAAC
Glulamic acid decarboxylase	Gad67	X57573	ATCTTGCTTCAGTAGCCTTTGC	TGTCTTCAAAAACACTTGTGGG
Neurokinin receptor 1	Nk1	J05097	CTGGAAAGAGGAGCCTTGTG	CTGAGACGGAAAGGAACAGC
Neurokinin receptor 2	Nk2	M31838	TTCTGCAGTGAGGAGCTGG	TTGGCTTTCAGAGGGCAC
Neurokinin receptor 3	Nk3	J05189	TTTGTGACCGCAGAAGAGC	TTGTGTGTGGGAAGTCAGTCA
Adenosine receptor 2a	Adora2a	M91214	TCTGACCAACAAAGCTGGC	TGGAAGGAAAGGCAGTAGTCA

* Brice A., et al (1989) J Neurosci. Res. 23:266-273

Example II

Reproducibility and Specificity of 3' Amplification of cDNA from Single Cells

The reproducibility of this technique was assessed when the expression of 20 genes was examined in duplicate on single lymphoblastoma cells. Duplicate PCR assays revealed almost identical expression profiles both between assays on a single cell and between different cells (Figure 3). The expression of 11 housekeeping genes was reproducible in all of these single cell expression profiles, demonstrating consistent amplification of cDNA from the transcripts of these genes. However four genes (JUND, EF1 CDC25B and CD19) were not found to be expressed in every cell tested. Two gene-specific assays, sucrase-isomaltase and CD2, were included as negative controls as they were not expected to be expressed in these cells since their mRNAs have only previously been observed in the gastrointestinal tract (Chadrasena G *et al* (1992), Cell Mol Biol, 38:243-254) and in populations of T-cells (Sewell W *et al* (1986) Proc Natl Acad Sci USA, 83:8718-8722), respectively. Another set of primers, designed to amplify an intronic sequence from a gene found in the Xq 2.5 region were used to detect genomic contamination. Since these three sets of primers were negative in all the cells examined (n=20), and expression of any gene was not detected without prior reverse transcription, false positives arising from amplification of genes not expressed in these cells

(SI and CD2) or genomic contamination, do not appear to complicate interpretation of results.

Figure 3 clearly demonstrates the reproducibility of the gene-specific assays following amplification from single cells, with each of 11 housekeeping gene assays being detectable in duplicate reactions on each of four lymphoblastoma cells. There was some apparent variability in the expression of four genes (JUND, EF1 CDC25B and CD19) studied in these cells. It is unclear at this stage however, whether this represents experimental variability, fluctuations in transcriptional activity within these cells or 'real' consistent differences in the expression of these genes in cells that we can only assume to be homogeneous. Such variability in gene expression has been encountered in other cell groups thought to be homogeneous (O'Dowd, D K & Smith M A (1996), Mol Neurobiol, 13:199-211). This analysis also demonstrates that the procedure does not lead to false positives due to either over amplification (as no signal was detected for genes (SI and CD2) known not to be expressed in these cells) or genomic contamination (as demonstrated by the lack of signal from intronic primers or from the gene-specific primers when the cell contents were not reverse transcribed prior to amplification).

The power of this technique lies in its potential to facilitate expression profiling of cells derived from complex cell populations, even when they form only a small proportion of the population as a whole, and in its ability to detect low abundance transcripts. Expression of the neurokinin (NK) receptors and the

adenosine A_{2a} receptor were investigate in single striatal cholinergic interneurons which constitute a small fraction of the total cellular mass of the striatum. Of the neurokinin receptors, the NK1 receptor is widely accepted as being expressed in these cells (Kawaguchi Y *et al* (1995), Trends in Neurosci, 18:527-535). The expression of this gene was examined as an example of an mRNA species expressed at far lower levels than housekeeping genes. In contrast, there is considerable controversy as to whether the A_{2a} receptor is expressed in cholinergic interneurons (Schiffman S N *et al* (1991), J Neurochem, 57:1062-1067; Fredholm B B & Svenningsson P (1998), Trends in Pharmacol Sci, 19:46-47, Richardson P J *et al* (1997), Trends in Pharmacol Sci, 18:338-344, Richardson P J *et al* (1998), Trends in Pharmacol Sci, 19:47-48), suggesting that the corresponding mRNA species may be present at low levels as suggested by one *in situ* hybridization study (Dixon A K *et al* (1996), Br J Pharmacol, 188:1461-1468) or not present at all, as suggested by other studies (Svenningsson P *et al* (1997) Neurosci, 80:1171-1185). Expression of the A_{2a} receptor in these cells has important implications for the mechanism of action of adenosine and because of the potential of the A_{2a} receptor to act as a target for novel drugs for the amelioration of Parkinson's Disease (Richardson P J *et al* (1997), Trends in Pharmacol Sci, 18:338-344). Using a patch pipette it was possible to harvest upwards of an estimated 40% of the cellular contents of these neurons (figure 4a). This was then subjected to the cDNA amplification procedure, followed by gene-specific PCR assays. Testing for the expression of

four housekeeping genes and the marker enzymes choline acetyltransferase and *Gad67*, was included in order to demonstrate the quality of collection and amplification, and to corroborate the cell lineage, respectively. NK1 receptor mRNA was detected in all the cholinergic neurons tested, although the NK2 and NK3 receptors were not, confirming that Substance P exerts its effects on these cells via the NK1 receptor (Bell M I *et al* (1998), Neurosci in press). Expression of the adenosine A_{2a} receptor mRNA was detected in 27% of the cholinergic neurons assayed (Figure 4b), a percentage close to that observed previously (Dixon A K *et al* (1996), Br J Pharmacol, 118:1461-1468). It is not clear whether the apparent heterogeneity in expression of this receptor is due to differences in the temporal expression of this gene (i.e. that all of these cells possess A_{2a} receptor protein, but only 27% actually express the gene at any one time), or to an absolute difference in gene expression within the striatal cholinergic interneuron population.

Example III

Extraction of Neuronal Contents

300 µm coronal slices from 14-28 day-old male Sprague Dawley rats containing the striatum were viewed with a Zeiss Axioskop microscope (Carl Zeiss Ltd, Welwyn Garden City, UK) fitted with a x64 water-immersion objective lens together with gradient contrast optics (Luigs and Neumann, Ratingen, Germany). Light in the infrared range (>740 nm) was used in conjunction with a contrast-enhancing Newvicon camera (Hamamatsu,

Hamamatsu City, Japan) to resolve individual neurones within slices (Stuart G J *et al* (1993), Pflügers Archiv, 423:511-518). The cytoplasm from large cells (>30 μm in one dimension) was gently aspirated under visual control into a patch-clamp recording electrode until at least 40% of the somatic cytoplasm had been collected. The electrode was then withdrawn from the cell to form an outside-out patch which prevented contamination when the electrode were forced into a microtube and reverse transcribed, subjected to 3' cDNA amplification, and 2.5% of the product used in each gene specific PCR reaction.

Example IV

Analysis of Complex Cell Systems: Expression Profiling of Neurons

Having shown that TPEA-PCR permits expression profiling of single cells, the analysis of a complex cell system can be made *in vivo*. Striatal cholinergic interneurons are readily identifiable in rat brain slices due to their large size (>30 μm diameter) when compared to surrounding cell types, which are predominately medium spiny neurons (<15 μm diameter). After electrophysical characterisation of the cells (Lee K *et al* (1997) J Neurochem, 69:1774-1776), cytoplasmic samples were taken using a patch pipette, reverse transcribed and the cDNA amplified. The expression of a variety of genes of was then investigated (two representative expression profiles of cholinergic interneurons are shown in Figure 4). The expression of four known housekeeping genes was demonstrated, confirming the integrity of sample

collection and the RNA. In addition, the expression of choline acetyltransferase (the acetylcholine synthesising enzyme) was observed in each sample, thus unequivocally confirming the cholinergic phenotype of the sampled neurons. In order to control against the possibility of contamination by the surrounding population of medium spiny neurons, which are known to express the adenosine A_{2a} receptor (Schiffman S N *et al* ((1991), J Neurochem, 57:1062-1067), the samples were assayed for the presence of the mRNA of glutamic acid decarboxylase (*Gad67*), which is highly expressed in medium spiny, but not cholinergic neurons. The *Gad67* primers amplify relevant mRNA from medium spiney neurons. Given the absence of any *Gad67* expression in cholinergic interneurons,the failure to see any amplification of *Gad67* mRNA in the sample demonstrates that samples were not contaminated with medium spiney neurons. All neurons tested for the expression of the neurokinin (NK) receptors were negative for NK2 and NK3 receptor mRNA, but positive for that of the NK1 receptor. 27% (7/26) of the cholinergic interneurones tested expressed the adenosine A_{2a} receptor.

Claims:

1. A process of reverse transcribing mRNA species present in a sample from an organism comprising:

- reverse transcribing the mRNA species using a first heeled primer, thereby to provide first strand cDNA species;

- synthesising second cDNA strands using a second heeled primer population, the nucleotide sequences of the non-heel portions of the second heeled primers being such that the reverse transcribed first strand cDNA species are capable of hybridising to at least one second primer.

2. A process of reverse transcribing and amplifying mRNA species present in a sample from an organism comprising:

- reverse transcribing the mRNA species using a first heeled primer, thereby to provide first strand cDNA species;

- synthesising second cDNA strands using a second heeled primer population, the nucleotide sequences of the non-heel portions of the second heeled primers being such that the reverse transcribed first strand cDNA species are capable of hybridising to at least one second primer.

- amplifying the resulting cDNA species.

3. A process as claimed in claim 2, wherein the amplification of the resulting cDNA species employs a third primer comprising at least a part of the heel portion of the first heeled primer and a fourth primer comprising at least part of the heel portion of the second heeled primer.

4. A process as claimed in any one of claims 1 to 3, wherein the second heeled primer population comprises primers differing by up to five nucleotide bases, the population comprising a number of primers in the range 1000 to 100,000 primers, preferably in the range 1024 to 65536 primers.

5. A process as claimed in claim 4, wherein the primers of the second heeled primer population each comprise a random sequence of nucleotides in the range of 5 to 8 nucleotides 3' to the heel and a further sequence of at least 5 nucleotides contiguous 3' therewith.

6. A process as claimed in claim 5, wherein the said further sequence of nucleotides is selected by sequence analysis of known sequences so as to promote the ability of the second heeled primer as a whole to hybridise to the transcribed cDNA species.

7. A process as claimed in claim 6, wherein the said further sequence of nucleotides comprises a number of nucleotides in the range 2 to 10 nucleotides.

8. A process as claimed in claim 7, wherein the said further sequence of nucleotides comprises a

number of nucleotides equivalent to the number of nucleotides in the random sequence of nucleotides.

9. A process as claimed in any preceding claim, wherein a second heeled primer hybridises on average once in every 1kb portion of a first strand cDNA species.

10. A process as claimed in any one of claims 5 to 9, wherein the said further sequence of nucleotides is:

CGAGA

11. A process as claimed in any one of claims 5 to 10, wherein the second heeled primer is:

CTGCATCTATCTAATGCTCCNNNNNCGAGA

wherein N is independently selected from C G T or A.

12. A process as claimed in any preceding claim, wherein the heel portion of the first and second heeled primers are selected so that they lack the ability to hybridise to mRNA or first strand cDNA respectively

13. A process as claimed in claim 12, wherein the heel portions are selected by an analysis of known nucleotide sequence information.

14. A process as claimed in claim 12 or claim 13, wherein the heel portions comprise sequences absent from the mRNA species in the sample.

15. A process as claimed in any one of claims 13 to 14, wherein the heel portions comprise sequences absent from the genome of the organism from which the sample is taken.

16. A process as claimed in any preceding claim, wherein the heel portions comprise a number of nucleotides in the range 15 to 30, preferably 18 to 22 nucleotides.

17. A process as claimed in any preceding claim, wherein the first heeled primer is an anchored primer comprising an oligo d(T) sequence.

18. A process as claimed in any preceding claim, wherein the fourth primer is the heel of the second heeled primer.

19. A process as claimed in any preceding claim, wherein the third primer is the heel of the first heeled primer.

20. A process as claimed in any one of claims 1 to 16, wherein the third primer is the same as the first heeled primer.

21. A process as claimed in any preceding claim, wherein the amplification of the resulting cDNA species comprises more than one round of amplification cycles.

22. A process as claimed in claim 21, wherein each further round of amplification comprises addition of

further second and third primers and amplification reagents.

23. A process as claimed in any preceding claim, wherein each round of amplification comprises 5 to 45 cycles, preferably 10 to 40 cycles.

24. A process as claimed in any one of claims 21 to 23, wherein the first round of amplification comprises a lesser number of cycles than any further rounds of amplification.

25. A process as claimed in any one of claims 21 to 24, wherein the amplification process is PCR.

26. A process as claimed in any one of claims 21 to 25, wherein each PCR cycle comprises X_1°C for Y_1 min; then X_2°C for Y_2 min; then X_3°C extension for Y_3 minute; then Y_4 min extension, wherein $X_1 > X_3 > X_2$ and $Y_4 > Y_3 > Y_2 > Y_1$ and wherein X_1 is in the range 90 to 94°C, X_2 is in the range 45 to 70°C, X_3 is in the range 65 to 75°C and Y_1, Y_2, Y_3 are in the range 15 seconds to 4 minutes and Y_4 is in the range 2.5 to 20 minutes.

27. A process as claimed in claim 26, wherein X_1 = 92 °C, X_2 = 60°C, X_3 = 72°C, Y_1 = 2.5 minutes, Y_2 = 1.5 minutes, Y_3 = 1 minute and Y_4 = 10 minutes.

28. A process as claimed in any preceding claim, wherein during the amplification of the cDNA strands, further primers are included for amplification of

contaminating nucleotide sequences suspected as being present in the sample.

29. A process as claimed in any preceding claim, wherein the sample from an organism is a cytoplasmic sample obtained from a single cell type, preferably from a single cell.

30. A process as claimed in claim 29, wherein the cytoplasm is obtained by lysis or aspiration of a cell or cells.

31. A process as claimed in claim 29 or claim 30, wherein cells are obtained by fluorescence activated cell sorting (FACS).

32. A method of reverse transcribing expressed gene sequences in a sample from an organism comprising reverse transcribing the mRNA in the sample using a first primer to produce first strand cDNA species, synthesising second cDNA strands using a population of second primers, wherein at least one second primer in the population hybridises to a given first strand cDNA species.

33. A method as claimed in claim 28, further comprising amplification of the resulting double stranded cDNA.

34. A method as claimed in claim 32 or claim 33 further comprising the features of any of claims 1 to 31.

35. A polynucleotide primer for reverse transcription of mRNA species comprising an oligo (dT) sequence and 5' thereto a polynucleotide heel sequence, wherein the heel sequence is substantially incapable of hybridisation to the mRNA species.

36. A primer as claimed in claim 35 being an anchored primer.

37. A primer as claimed in claim 35 or claim 36 further comprising the features of any of claims 12 to 16.

38. A polynucleotide primer population for synthesis of second strand cDNA species from first strand cDNA species, wherein at least one primer in the population is capable of hybridising to a given first strand of cDNA.

39. A primer population as claimed in claim 38, wherein at least one primer in the population is capable of hybridising at least once in any given 1kb of first strand cDNA.

40. A primer population as claimed in claim 38 or claim 39, wherein the primers further comprise the features of any of claims 4 to 16.

41. Polynucleotide primers for amplification of cDNA comprising a primer as claimed in any one of claims 35 to 37 and a primer comprising at least a portion of the heel portion of the primers claimed in claim 40 or claim 40 when dependent on any of claims 4 to 16.

42. The use of a polynucleotide comprising an oligo d(T) sequence and a heel sequence 5' thereto for the reverse transcription of mRNA species in a sample.

43. The use as claimed in claim 42, wherein the polynucleotide further comprises the features of any of claims 35 to 37.

44. The use of a polynucleotide primer population of any of claims 38 to 40 for the synthesis of second strand cDNA from a population of first strand cDNA species.

45. A cDNA library preparation obtainable by a process of any of claims 1 to 31 or a method of any of claims 32 to 34, said library comprising substantially all cDNA species corresponding to genes expressed by a single cell, cell type or tissue.

46. A kit for the production of cDNA from mRNA in a sample from an organism comprising a primer of any of claims 35 to 37 and a primer population of any of claims 38 to 40.

47. A kit as claimed in claim 46, further comprising at least one further primer for achieving amplification of the cDNA.

48. A kit as claimed in claim 46, further comprising a primer pair for amplification of the cDNA.

Fig.1.

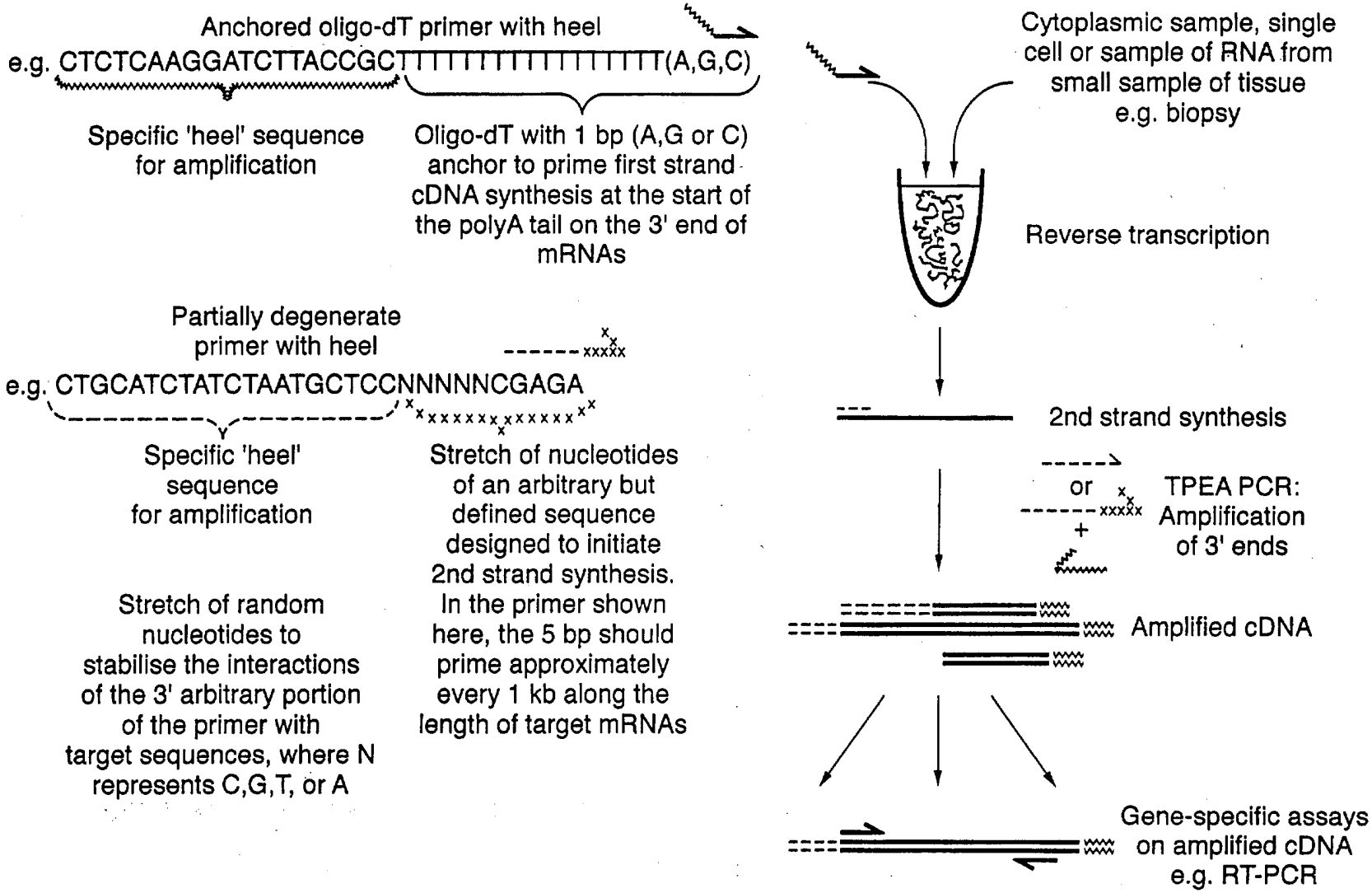


Fig.2.

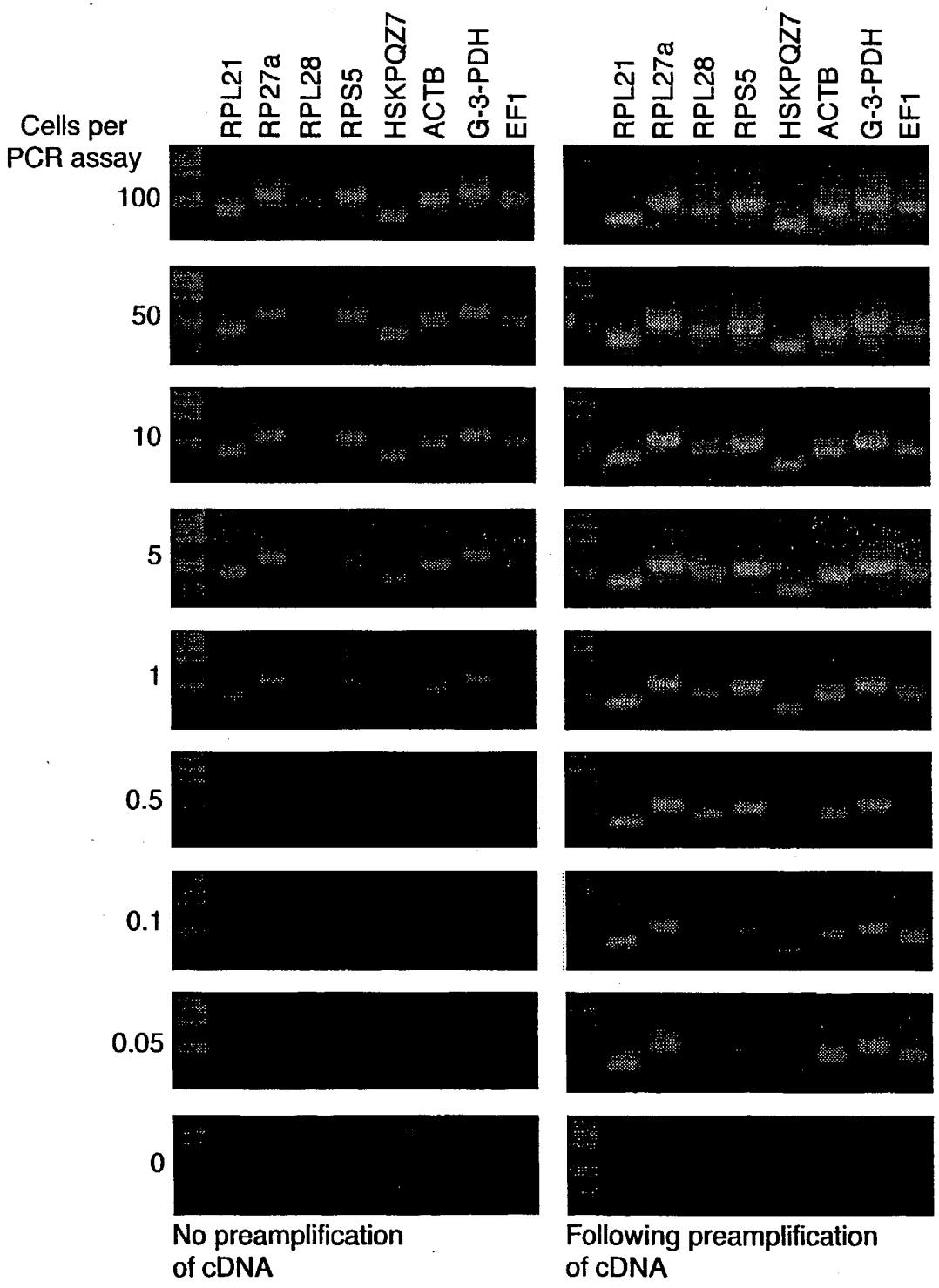

RPL21
RP27a
RPL28
RPS5
HSKPQZ7
ACTB
G-3-PDH
EF1
RPL21
RPL27a
RPL28
RPS5
HSKPQZ7
ACTB
G-3-PDH
EF1
Cells per PCR assay
100
50
10
5
1
0.5
0.1
0.05
0
No preamplification of cDNA
Following preamplification of cDNA

Fig.3.

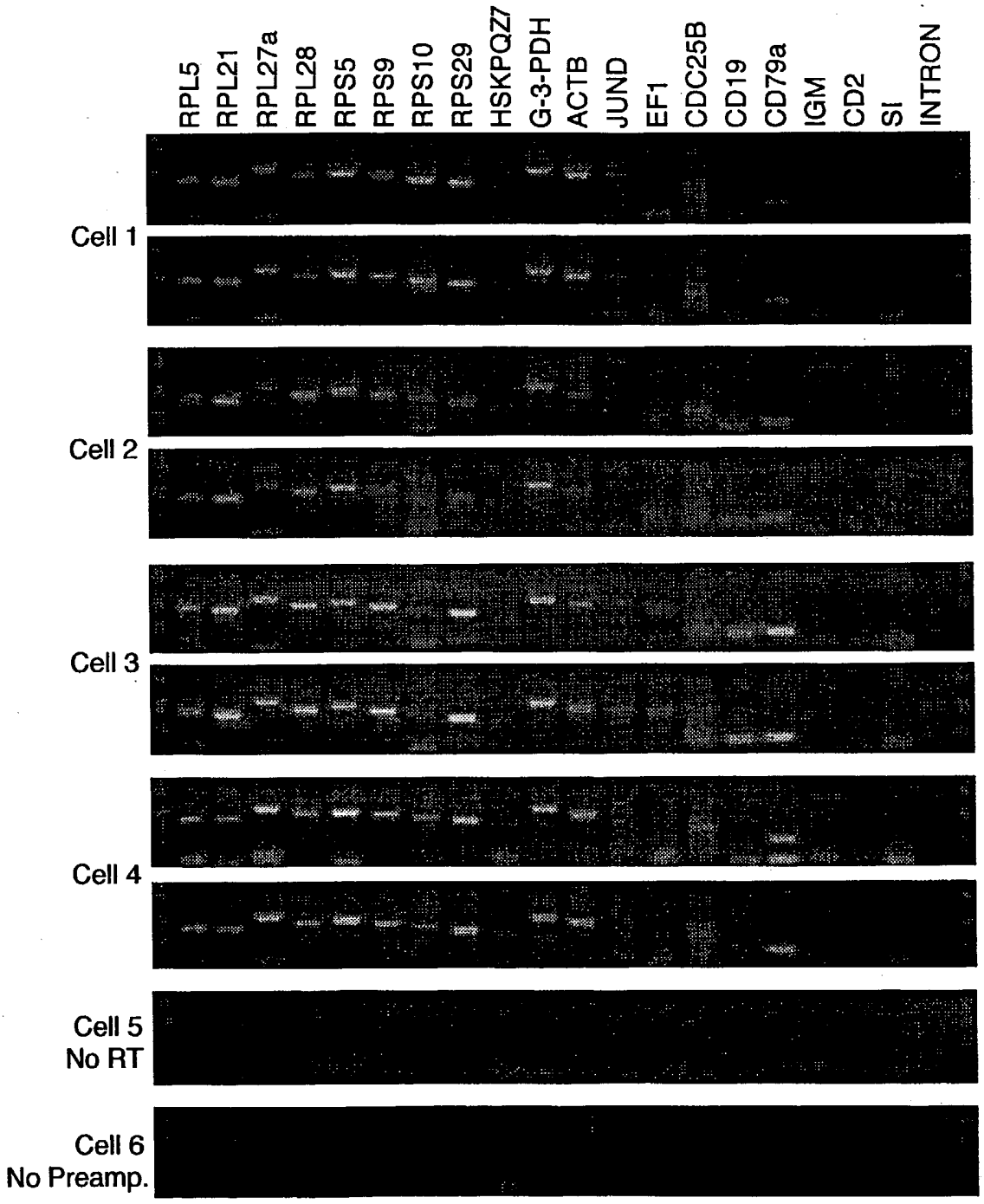
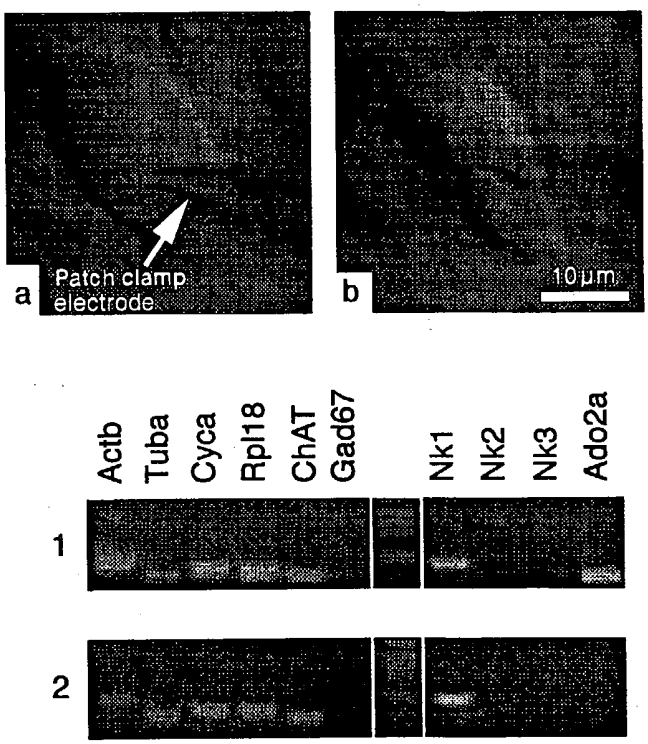

RPL5
RPL21
RPL27a
RPL28
RPS5
RPS9
RPS10
RPS29
HSKPQZ7
G-3-PDH
ACTB
JUND
EF1
CDC25B
CD19
CD79a
IGM
CD2
SI
INTRON
Cell 1
Cell 2
Cell 3
Cell 4
Cell 5
No RT
Cell 6
No Preamp.

Fig.4.


a
Patch clamp electrode
b
10 µm
Actb
Tuba
Cyca
Rpl18
ChAT
Gad67
Nk1
Nk2
Nk3
Ado2a
1
2